DOCUMENT RESUME

ED 409 346 TM 026 829

AUTHOR Swaak, Janine; And Others
TITLE Simulation Based Discovery Environments and Acquisition, the Features, and Assessment of Intuitive Knowledge.
PUB DATE 28 Mar 97
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS College Students; *Discovery Learning; *Educational Assessment; Evaluation Methods; Foreign Countries; Higher Education; *Intuition; *Simulation; *Test Construction; Test Items; Test Use; Test Validity
IDENTIFIERS *Knowledge Acquisition

ABSTRACT

A study was conducted to develop a test that is able to capture knowledge of an intuitive nature, such as that acquired through discovery learning. The proposed test format is called the "what-if test." Test items in this format consist of the presentation of a situation. A change in the situation is introduced, and learners have to select as quickly as possible, from a number of predefined situations, the resulting situation. Performance of students is measured in terms of correctness and latency data. For validation purposes, performance on the what-if test is related to scores on a definitional knowledge test and several aspects of interaction with the discovery environment. This paper presents three main characteristics of simulation-based discovery environments (richness, low transparency, and interaction) and explains how they relate to the acquisition and features of intuitive knowledge. Three empirical studies in which a first validation has been performed are reported. These were investigations of learning through simulation about: (1) elastic collisions by 46 college students; (2) one-dimensional oscillatory motion by 28 college students; and (3) electrical circuits by 41 subjects. An overall conclusion is that tests constructed with what-if items were able to tap an improvement in learning. The approach seems promising for new types of knowledge tests in relation to new types of technology-supported learning. (Contains 4 figures, 6 tables, and 21 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 409 346

Simulation based discovery environments and acquisition, the features, and assessment of intuitive knowledge

Paper presented at AERA, Chicago, Illinois, March 28, 1997

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

JANINE SWAAK, TON DE JONG & WOUTER R. VAN JOOLINGEN
Faculty of Educational Science and Technology
University of Twente, PO Box 217
7500 AE Enschede, The Netherlands
E-mail: swaak@edte.utwente.nl

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Janine Swaak

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Abstract: Learning in discovery environments, such as simulations, is supposed to lead to acquisition of knowledge that is qualitatively different from knowledge that is acquired in more traditional instructional situations. In traditional instruction learning usually involves an explicit transfer of a body of knowledge, with an emphasis on the reproduction of the taught knowledge. In discovery environments, the learning processes invoked may be less explicit, and the resulting knowledge may therefore also be of a less explicit character. The knowledge that is acquired through discovery is frequently characterised with labels as 'intuitive', 'deep', 'qualitative' etc. The problem is that no good tests for this type of knowledge are available so that the results of discovery learning remain unclear. The objective of our research is to develop a test that is able to capture knowledge of an 'intuitive character'. The proposed format is what we call the 'WHAT-IF test', of which the essential characteristics are that items consist of the presentation of a situation, a change in the situation is introduced, and learners have to select from a number of pre-defined situations, the resulting situation as quickly as possible. Performance of students is measured in terms of correctness and latency data. For validation purposes performance on the WHAT-IF test is related to scores on a definitional knowledge test and several aspects of the interaction with the discovery environment. In the paper we present three main characteristics of simulation based discovery environments, and explain how they relate to the acquisition and features of intuitive knowledge. This makes up the theoretical background that is used for developing the WHAT-IF format. Then we shortly describe the requirements we pose to this type of test format, and report on three empirical study studies in which a first validation was performed.

TM026829

1. Introduction

Learning in discovery environments, such as simulations, is supposed to lead to acquisition of knowledge that is qualitatively different from knowledge that is acquired in more traditional instructional situations. In traditional instruction, learning usually involves an explicit transfer of a body of knowledge, with an emphasis on the analytical aspects of the domain and on the reproduction of the taught knowledge. In discovery environments, the learning processes invoked may be less explicit, and the resulting knowledge may therefore also be of a less explicit character. Currently, however, there appears to be a mismatch between the context of learning, the qualities of the knowledge acquired and the methods used to assess these qualities of knowledge (see also Snow, 1989). Thomas and Hooper (1991), for example, classified and analysed simulation studies according to the instructional function for which the simulation was used. The conclusion most relevant for the present work is that in the authors' words: "the effects of simulations are not revealed by tests of knowledge (....)" (p. 479). They further conclude that simulations can best serve as "experiencing programs" which give students the opportunity "to gain an intuitive understanding of the learning goal" (p. 499). Thomas and Hooper do, however, not indicate what they mean by "tests of knowledge", why and how they think an intuitive understanding is gained, and how the intuitive understanding can be assessed.

BEST COPY AVAILABLE

The objective of our work is to have a closer look at the relations between the nature of simulation-based discovery environments, the processes elicited, the resulting knowledge, and a method to tap the acquired knowledge.

2. Characteristics of simulation based discovery environments

In our studies we work with simulations that are developed with the SMISLE[1] system. The environments of the SMISLE project carry a number of features (e.g., De Jong et al., 1992; De Jong & Van Joolingen, 1995). At the heart of such an environment is always the simulation in which the subject matter is simulated. In the simulation learners can manipulate variables the way they choose and see the consequences of their manipulations in dynamic outputs. Furthermore, the simulation environment is enriched with support measures added to the simulation The support measures that can be created with SMISLE include explanations, assignments, and model progression. The explanations consist of explanations of variables of the domain. Students can at own will, choose to look up an explanation in which case a description of the variable is given. Assignments are small exercises that aim to help students structure their learning and to point them to important phenomena in the domain. Among them are investigation assignments which ask the student to investigate the relation between to variables of the domain by manipulating the variables in the simulation, and prediction assignments in which the student is asked to predict the relation between two variables in the domain by choosing the predicted value of a variable in the simulation. Model progression means that the domain is offered in small subsequent steps, where in every step new variables are added to the model. Also model progression is offered for structuring the learning process (see White & Frederiksen, 1990).

It is argued that these types of learning environments can be described as *'rich'* environments. By saying this we mean that a) a high amount of information can be extracted by the learner, that b) this information can be obtained in several ways, and that c) the information is usually displayed in more than one representation; a dynamic, graphical representation of the output is generally present next to animations and to numerical outputs.

A second feature of simulation-based environments concerns the relatively *low transparency* of the learning environment (as compared to text books or hypertexts etc.). The less transparent the discovery environment, the less a 'direct view' on the variables and relations is given to the learners (de Jong & van Joolingen, 1997). Consequently, the more is to be inferred or extrapolated. Simulation environments generally have a low transparency, only by adding instructional measures such as explanations, the transparency can be enlarged a little.

The third characteristic of the learning context of this work involves the *interaction* aspect. The learning session entails an interaction with a simulation based environment. Learners are not supposed to passively absorb information on the domain from the computer screen, but are expected to perform several kinds of actions, next to reading and thinking, to make up their own 'meaningful' learning session. Thus, the learning with the discovery environment may accurately be described as an active 'experience'.

We argue that the characteristics of the learning context have a considerable influence on the leaning processes displayed. For simulation-based discovery learning the nature of learning can be described as experiential. The features of the learning processes, in turn, will have an impact on the learning products. The knowledge aimed at might be described as 'intuitive' knowledge. In the following section we will see that both literature on intuitive knowledge and research on interacting with complex simulations systems point to this direction.

[1] The SMISLE System is an authoring environment for developing Multimedia Integrated Simulation Learning Environments. The SMISLE project and its' follow-up the SERVIVE project were/are partly sponsored by the EC (DG XIII) in its Telematics programme. In the course of the SMISLE and the SERVIVE projects, several applications, this is integrated simulation learning environments, have been developed and evaluated. The aim was/is to estimate the influence of the supportive measures (assignments, explanations, and model progression) on the qualities of the learning processes (cognitive processes and system operations), the height of the cognitive load experienced, and on the qualities of the learning products. In this paper we concentrate on the qualities of the product as measured with WHAT-IF tests. Part of present work is reported in *Studies in Educational Evaluation, 22 (4),* 341-362.

3. Acquisition of intuitive knowledge

As an example from the literature on intuition, we take the work of Fischbein. Fischbein (1987) states that intuitions "can never be produced by mere verbal learning (.....) but that they "only can be attained as an effect of *direct, experiential involvement* of the subject in a practical or mental activity" (p. 95). In addition, Fischbein argues that intuitions are implicit, that they are based on complex selection and *inference processes*, which are, to a large extent, believed to be *unconscious* to the individual. Likewise, Westcott writes that it seems that intuition "occurs when an individual reaches a conclusion on the basis of less explicit information that is ordinarily required to reach that conclusion" (1968, p. 97; cited in Fischbein, p. 50).

Broadbent and colleagues studied implicit knowledge by looking at the performance of students controlling *complex dynamic systems* and related implicit knowledge to implicit learning. They have speculated that "if learners follow an implicit learning mode they probably build up in parallel large number of contingencies rather than following through and evaluating one or more hypotheses" (p.271). They further say that the contingencies built up by implicit learning could possibly be represented as a list of procedures or as Broadband, Fitzgerald and Broadbent. (1986) have suggested as form of look-up table. Berry and Broadbent (1988) conjectured that "in either case a particular set of circumstances or conditions would give rise to a particular response. Whatever the precise nature of the representation, it is likely that some general overall 'pattern-matching' process plays a critical role." (Berry & Broadbent 1988, p.271).

Norman (1993) reasons that some environments lead one more toward an 'experiential' mode of learning while in other contexts learning can be characterised as 'reflective'. We argue that the learning environments referred to in this work have some of the important features that foster an experiential way of learning (for a discussion see Swaak & de Jong, 1996).

To recapitulate, we stated that the characteristics of the learning context has a considerable influence on the leaning processes displayed, and that the nature of discovery learning with simulations here consists, to a substantial extent, of experiential learning. The features of the learning processes, in turn, have an impact on the learning products. In this work we are primarily interested in the intuitive quality of knowledge, because we postulate that this is the quality most favoured by experiential learning processes elicited in the interaction with a computer simulation.

4. Features of intuitive knowledge

Despite the under-representation of serious efforts to assess intuitive knowledge, literature on intuitive knowledge (e.g., Fischbein, 1987) has provided us with at least three features of intuitive knowledge The first entails that if intuitive knowledge is based on information that is not explicitly stated and/or if processes involved in intuitive knowledge are to a large extent unconscious, then intuitive knowledge should be *hard to verbalise*.

Secondly, Fischbein writes about *self-evidence* and *immediacy* of intuitions. This means that an intuition "appears subjectively to the individual as directly acceptable" (p. 200). Fischbein calls "visualisation" the main factor responsible for the effect of immediacy (p. 103). *Visualisation* is the third feature of intuitive knowledge. According to the Fischbein visualisation may or may not be mediated by an external representation. The author, moreover, recalls that "what one cannot imagine visually is difficult to realise mentally" (p. 103). Though visualisation is critical, Fischbein terms an intuition a theory, not just perception. Yet, he calls an intuition "the analog of perception at the symbolic level".

To sum up, low verbalisability, immediacy, and visualisation are features frequently mentioned in relation to intuitive quality of knowledge.

A certain coherence can be indicated between on the one hand, the characteristics of simulation based discovery environments, and on the other hand the acquisition and features of intuitive knowledge (see Figure 1 for an overview). Several questions remain unanswered, however. So far, there is no agreement on the exact nature of the processes involved in the acquisition of the intuitive knowledge. Even more remains unclear about the precise representation of intuitive conceptual knowledge.

characteristics of simulation based discovery environments	*acquisition of intuitive knowledge*	*features of intuitive knowledge*
perceptually 'rich'	direct experiential involvement	hard to verbalise
low transparency	extrapolation, inference	directly available or not available at all
active experience involved	in rich dynamic environments	perception, visualisation
	'quick perception of meaningful situations'	

Figure 1. *Overview of concepts discussed with regard to characteristics of simulation based discovery environments, the acquisition of intuitive knowledge, and features of intuitive knowledge*

However, most researchers agree that, whatever the exact nature of the processes involved in the acquisition and whatever the precise representation of intuitive conceptual knowledge, the processes involved in the manifestation of the intuitive quality of knowledge can be described as a 'the quick perception of meaningful situations'.

5. Selection of item format and test procedure

Based on the notion that 'quick perception of meaningful situations' is a critical feature related to the intuitive quality of knowledge, we designed, what we call, the WHAT-IF item format.

In the WHAT-IF task conceptual knowledge is presented in the form of conditions, actions, and predictions. The conditions and predictions are the states or situations in which the system can be involved, and can be displayed in a drawing of the system, or in a graph or tabular form, accompanied with a minimal amount of text. In addition predictions concerning one variable can be presented by one word or number. The action, or the change of a variable within the system, is presented in text and/or graph. The items of the task are kept as simple as the domains permitted. Items can have a qualitative or a quantitative character, but in the case of quantitative items, no calculations are needed to give a correct answer. Knowledge about the specific nature of the relation between the variables is sufficient. A complete justification of the item format, is given in Swaak and de Jong (1996).

An important aspect of the procedure is that learners are not only asked to give the correct answer, but they are also required to do so as quickly as they possibly can. With these features, the WHAT-IF format and procedure are clearly different from the format and procedures used in traditional multiple-choice tests.

In the next sections we present three study studies in which a first evaluation of the new format was performed. Example items are displayed in Figure 2, Figure 3, and Figure 4. The items had a three-answer format. The task was computer administered. The moment learners had clicked with the mouse on the answer of their choice, the item disappeared from the screen and the next item popped-up. Learners could not go back to previous answered items.

6. Discussion of first three studies

6.1 General procedure

The studies followed a pre-test post-test design[2], and parallel versions of WHAT-IF tests were applied as pre-and as post-tests to every learner. In order to gain insight in the validity of the WHAT-IF test (i.e., does the test tap conceptual knowledge of an intuitive quality, and not, for example, conceptual knowledge of a declarative quality) the learners were also administered a definitional knowledge test.

The definitional knowledge test, consisted of three-answer items and aimed to measure conceptual knowledge of 'isolated' facts and definitions of the domain. The same definitional test was given both

[2] The learners in the study studies were divided across experimental conditions (in Study 1 we created three, in Studies 2 & 3, two groups) which differed with respect to the instructional measures added to the simulation (Studies 1 & 2), or on the degree to which learners were given freedom to use the instructional measures (Study 3). In this paper only analyses that were performed across the conditions are reported.

as pre- and as post-test. The correctness scores on the definitional tests were compared with the correctness scores and the item completion times of the WHAT-IF tests.
Apart from gathering data on knowledge acquired, we registered all the actions learners made while interacting with the simulation. This provided us with data on the use of the simulation and the supportive measures that were present. These process measures were taken into account in the validation of the WHAT-IF format. The correctness scores and the completion times of the WHAT-IF tests were correlated with the frequency of use of the several types of support.

The first learning environment deals with the domain of elastic collisions and is called Collision, in the second environment harmonic oscillations are treated, the system is called SETCOM, the third study evaluates Circuit, a simulation environment on the domain of electrical circuits. (A full description of Study 1 can be found in De Jong et al., 1995. Study 2 is fully described in van Joolingen, van der Hulst, Swaak, & de Jong, 1995. Study 3 is described in Swaak, Blokhuis, Guiterrez, & López, in preparation.)

7. Study 1: Collision

7.1 Domain of the study and the what-if items created

The central topic of the learning environment Collision involves elastic collisions. More specifically, the rules behind colliding objects of different mass and velocity, and the concept of 'centre of mass' are addressed. Furthermore, both kinetic and potential energy approaches are part of the simulation environment. For the evaluation of Collision two parallel what-if tests each consisting of 37 items were developed. Two example items are displayed in Figure 2.

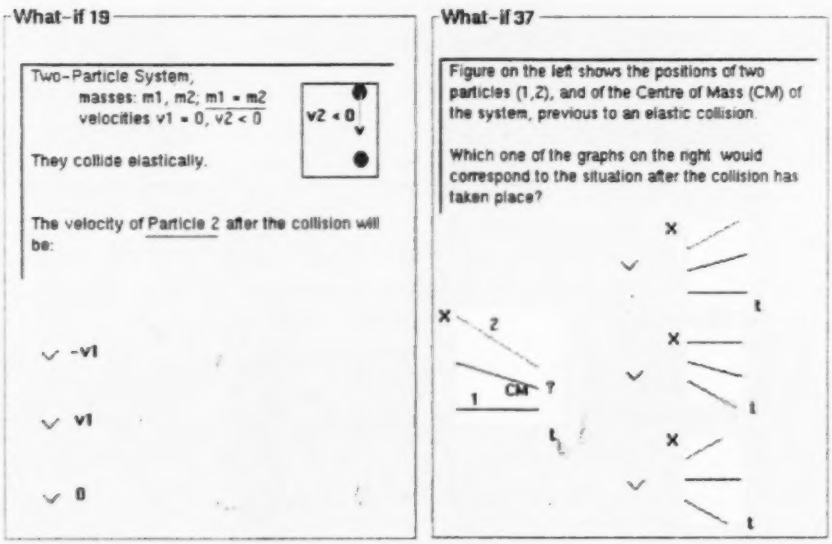
What-if 19
Two-Particle System,
masses: m1, m2; m1 = m2
velocities v1 = 0, v2 < 0
v2 < 0
They collide elastically.
The velocity of Particle 2 after the collision will be:
-v1
v1
0
What-if 37
Figure on the left shows the positions of two particles (1,2), and of the Centre of Mass (CM) of the system, previous to an elastic collision.
Which one of the graphs on the right would correspond to the situation after the collision has taken place?
X
t
2
CM
1

Figure 2. *Two examples of WHAT-IF test items used in the Collisions study. In the item on the left predictions are represented by one number/symbol. In the right hand side item the predictions are displayed by graphs (the items in the figure are translated from Spanish).*

7.2 Subjects

Students (N = 46) were selected from the populations of students from the first year university level and came from the Mathematics/computer-science and Biology departments. Both types of students have comparable backgrounds in high school, but the students from Mathematics/Computer-science have passed through a selection process previous to be allowed to study Mathematics/Computer-science.

7.3 Procedure

The Collision WHAT-IF pre- and post-tests were administered to 46 learners before and after they worked, for about an hour, with the Collision simulation environment. Before the learners started to work on the WHAT-IF tests they completed a definitional knowledge test, which consisted of 36 three-

BEST COPY AVAILABLE

answer items. The same definitional test was administered both as pre- and as post-test version. The session ended after completion of the WHAT-IF post-test.

7.4 WHAT-IF test results

The average number correctly answered items on the WHAT-IF pre-test was 19 with an SD of 4. On the WHAT-IF post-test the number correct scores had a mean of 23 with a SD of 5. Five of the 46 subjects answered less items correctly on the post-test as compared to the pre-test. The range was 1 to 5 items less. The average time to complete a WHAT-IF pre-test item was 34 seconds with a SD of 7.9. For the WHAT-IF post-test the average item completion time was 23 seconds and the SD was 7.0. One subject took longer on the post-test than on the pre-test. For this subject, on the average 7 seconds more were needed for the post-test items.

No trade-off between correctness and speed was found. The correlations between answer time and correctness had a value of $r = -.17$, $p > .10$ when computed within students across the WHAT-IF pre-test items, a value of $r = -.23$, $p > .10$ when computed within students across the WHAT-IF post-test items, a value of $r = .02$, $p > .10$ when computed within WHAT-IF pre-test items across students, and finally a value of $r = -.27$, $p > .10$ when computed within WHAT-IF post-test items across students.

Repeated measurement analyses on the WHAT-IF test scores showed both a significant within-subject effect of number of correct items ($F_{1, 43} = 36.34$, $p < .05$) and a significant within-subject effect of the WHAT-IF completion times ($F_{1, 43} = 125.09$, $p < .05$).

7.5 Relationship between the WHAT-IF test scores and definitional test score

The first kind of validation data consisted of the results of the definitional knowledge test. The definitional knowledge test was given in the same form as pre- and as post-test. It consisted of 36 multiple choice items with 3 alternative answers each. A repeated measurement analysis showed a significant within-subject effect of number of correct items ($F_{1, 43} = 37.56$, $p < .05$). In Table 1 the correlations between the WHAT-IF post-test scores and the definitional test correctness scores are given.

Table 1. *Correlations between the WHAT-IF post-test scores and the definitional knowledge test scores*

	definitional post-test	WHAT-IF post-test completion times
WHAT-IF post-test correctness	.80 ($p < .01$)	.02 ($p > .10$)
WHAT-IF post-test completion times	-.11 ($p < .10$)	

The pattern that emerges from this analysis is that we find two clusters. The first one consists of the definitional test and WHAT-IF correctness, the second one is the WHAT-IF time aspect. The correlation between the definitional post-test and the WHAT-IF post-test correctness score is considerable. However, the correlation between the gain in definitional test score and gain in WHAT-IF correctness is .16 and does not reach significance, $p > .10$.

7.6 Relationship between the WHAT-IF test scores and several process measures

We registered all the actions learners made while interacting with the simulation. This provided us with data on the use of the simulation and the supportive measures that were present. Due to technical issues resulting in loss of logfiles, it was not possible to collect the interaction data of all the learners. In the subsequent analyses the complete interaction data of 34 subjects were used. All data presented on interaction behaviour should be seen in this context.

As all the learners used the support measures available to them to a considerable extent, not much variation between subjects existed. However, there are some interesting relations found between use of assignments and scores on the WHAT-IF post-test. Table 2 displays the WHAT-IF post-test scores and number of explanations and assignments used.

Table 2. *Correlations between the WHAT-IF post-test scores and number of explanations and assignments used*

post-test correctness scores	number of explanations	number of assignments
WHAT-IF correctness scores	.15 ($p > .10$)	.51 ($p < .05$)
WHAT-IF completion times	.11 ($p > .10$)	-.35 ($p < .10$)

Table 2 shows a significant correlation between the number of assignments and the correctness scores of the WHAT-IF post-test, and non-significant (at the $\alpha < .05$ level), but moderate negative correlation between the numbers of assignments and the completion times of the WHAT-IF post-test. These figures indicate that a higher number of assignments is associated with a higher correctness score on the WHAT-IF post-test, and suggest that the more assignments a learner has utilised, the lower his/her completion times of the WHAT-IF post-test are.

7.7 Discussion

The overall picture that emerges is quite satisfactory. A knowledge gain is spotted, and established as both an increase in correct WHAT-IF items and a decrease in WHAT-IF test completion times. Furthermore, the items seem neither too difficult nor too easy. Importantly, no trade-off is found between the correctness and the completion time of items. There even seems an indication of the reverse: the quicker an item is answered, the higher the chance it is correct. The suggested direction is clearly in line with our ideas on the intuitive quality of conceptual knowledge as the 'quick perception of meaningful situations'.

However, not only a significant gain on the WHAT-IF tests is found, but also on the definitional tests, and more essentially, a strong correlation is found between the correctness scores of the WHAT-IF post-test and the scores of the definitional post-test. The high correlation makes it more difficult to maintain that the tests measure different aspects and qualities of the domain knowledge. Yet, if we look at the gain scores we can say that a gain in intuitive knowledge does not automatically yield a gain in definitional knowledge. Also, the correlation between the completion times of the WHAT-IF post-test and the scores on the definitional test is not found significant.

Finally, we mention the correlation between the number of assignments used and the WHAT-IF post-test scores. The moderate correlations may suggest that the more assignments learners have utilised, the higher their WHAT-IF post-test correctness scores and the lower their completion times of the WHAT-IF post-test. These figures nicely correspond with our conceptions that assignments would invite students to a more experiential mode of learning and have a beneficial effect on the intuitive knowledge of the students.

8. Study 2: SETCOM

8.1 Domain of the study and the WHAT-IF items created

The learning environment SETCOM deals with the principles of one-dimensional oscillatory motion. It incorporates three kinds of oscillatory motion: free oscillatory motion without friction, damped motion, and forced oscillatory motion. Moreover the concepts of 'sub- and supercritical damping', 'resonance' and the characteristic equation are addressed. For the evaluation of SETCOM two parallel WHAT-IF tests consisting each of 34 items were developed.

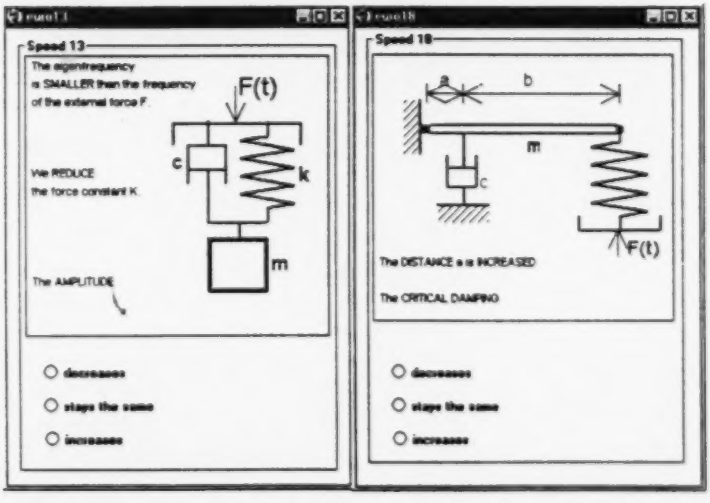

Figure 3. *Two examples of WHAT-IF test items used in the SETCOM study.*

The tests consisted of both qualitative (21 items) and quantitative items (13). Two examples of WHAT-IF items are displayed in Figure 3

8.2 Subjects

Twenty-eight subjects participated in the study. They were undergraduate students in Mathematics/Computer-science, Chemistry, and Psychology. All subjects had had physics including classical mechanics at A level, only the Mathematics/Computer-science students had a background in the theory of complex numbers. Most subjects were *not* familiar with the domain of oscillatory motion.

8.3 Procedure

The procedure in this study is the same as in the previous study, except that here the 28 learners worked for about two hours with the SETCOM simulation environment.

8.4 WHAT-IF test results

The average number correctly answered items on the WHAT-IF pre-test was 11 with an SD of 4. On the WHAT-IF post-test the number correct scores had a mean of 16 with a SD of 3. Five of the 28 subjects scored worse on the WHAT-IF post-test compared to the pre-test. The range was 1 to 9 items less correct. The average time to complete a WHAT-IF pre-test item was 25 seconds with a SD of 6.6. For the WHAT-IF post-test the average item completion time was 20 seconds, and the SD was 6.8. Eight subjects needed more time for the post-test as opposed to the pre-test. The 'extra' time taken ranged from .25 to 7.5 seconds for average item completion time.

We did not find a trade-off between correctness and speed. The correlations found between answer time and correctness had a value of $r = .00$, when computed within students across the WHAT-IF pre-test items, a value of $r = .00$, when computed within students across the WHAT-IF post-test items, a value of $r = -.16$, $p > .10$ when computed within WHAT-IF pre-test items across students, and finally a value of $r = -.19$, $p > .10$ when computed within WHAT-IF post-test items across students.

Repeated measurement analyses were performed on the WHAT-IF test scores. The analyses showed both a significant within-subject effect of number of correct items ($F_{1, 26} = 18.34$, $p < .01$) and a significant within-subject effect of the test completion times ($F_{1, 26} = 12.93$, $p < .01$).

8.5 Relationship between the WHAT-IF test scores and definitional test scores

The definitional knowledge test was given in the same form as pre- and as post-test. It consisted of 30 multiple choice items with 3 alternative answers each. A repeated measurement analysis on the definitional test-scores showed *no* significant within-subject effect of number of correct items ($F_{1, 25} = 1.77$, $p > .10$). Table 3 displays the correlations between the post-test scores of the WHAT-IF tests and the definitional knowledge test scores.

Table 3. *Correlations between the post-test scores of the WHAT-IF tests and the correctness scores of the definitional knowledge test*

	definitional post-test scores	**WHAT-IF post-test completion times**
WHAT-IF post-test correctness	.42 ($p < .05$)	-.01($p > .10$)
WHAT-IF post-test completion times	.19 ($p > .10$)	

Here, we see the same pattern as in the previous study. Again, the correlation between the definitional post-test and the WHAT-IF post-test correctness score is considerable. Also here, the correlation between the gain in definitional test score and gain in WHAT-IF correctness is less extensive, .31 and does not reach significance, $p > .10$.

8.6 Relationship between the WHAT-IF test scores and several process measures

All the learners used the support measures available to them to a considerable extent and therefore variation between subjects was small. Still, some correlations are found that are worth mentioning. Table 4 displays the correlations.

Table 4. *Correlations between the WHAT-IF post-test scores and number of assignments and explanations used*

post-test scores	number of assignments	number of explanations
WHAT-IF correctness scores	.00 ($p > .10$)	-.19 ($p > .10$)
WHAT-IF completion times	-.38 ($p < .10$)	.38 ($p < .05$)

The correlation reaching a level of $\alpha < .05$, indicates that the higher number of explanations are used, the higher the WHAT-IF post-test completion times are. The negative correlation of -.38 suggests that the more assignments are used, the lower the WHAT-IF post-test completion times are. This same negative correlation was found in the previous study.

8.7 Discussion

The general impression of this study is that the performance on the post-test measures is quite low and that the variance in post-test scores is substantial. More specifically, the correctness scores on the WHAT-IF pre-test are at chance level, and the correctness scores on the WHAT-IF post-test are still near chance level.

Nevertheless, both a significant increase in correctness scores and a significant decrease in completion times of the WHAT-IF tests are established, and furthermore, no trade-off between correctness and completion times are found. Like in the Collision study, there even seems an indication of the reverse which fits well with our hypotheses.

In contrast to the gain in WHAT-IF test scores, *no* increases in definitional knowledge test scores are attained. Another interesting finding is the relation found between the number of assignments used and the results on the WHAT-IF post-test. A moderate negative – not significant at the $\alpha < .05$ level – correlation is found between the number of assignments and the completion times of the WHAT-IF post-test. So, this may suggest that the more assignments the learners utilised, the quicker they completed the WHAT-IF post-tests. The reverse is established for the number of explanations: the more explanations are looked up, the longer the completion times of the WHAT-IF post-test. The rationale put forward here is that explanations foster a reflective mode of learning, resulting in higher response times of WHAT-IF items, whereas assignments invite learners to follow a experiential learning mode which is assumed to improve the intuitive quality of knowledge, as reflected in lower response times of WHAT-IF items.

9. Study 3: CIRCUIT

9.1 Domain of the study and the WHAT-IF items created

In this application the behaviour of current and voltage sources in electrical circuits is simulated. Conceptions of circuits with one ideal current source and one or more resistors are covered, in both serial, parallel and combined circuits. In addition, the law of Ohm, the current law of Kirchhoff, and the voltage law of Kirchhoff are included in the domain presentation. For the evaluation of CIRCUIT two parallel WHAT-IF tests consisting each of 24 items were used. The tests consisted of only qualitative items. Two examples of WHAT-IF items are displayed in Figure 4

9.2 Subjects

Forty-one subjects participated in the study. They were first and second year students from middle vocational technical training. The students had some basic knowledge of Electrical circuits and ample computer experience. Earlier in the study year, the first year students had followed an introductory course on Electrical circuits, for the second year students a year had passed since they followed this course.

9.3 Procedure

The procedure in this study is the same as in the previous studies, here the 41 learners worked one hour and ten minutes with the CIRCUIT simulation environment.

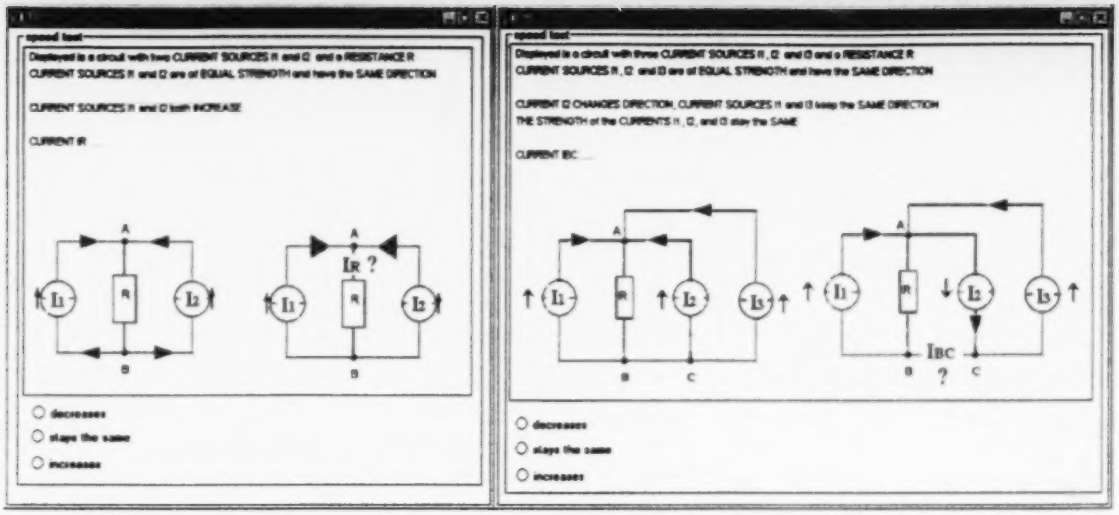

Figure 4. *Two examples of WHAT-IF test items used in the CIRCUIT study.*

9.4 WHAT-IF test results

The average number correctly answered items on the WHAT-IF pre-test was 12 with an SD of 4. On the WHAT-IF post-test the number correct scores had a mean of 14 with a SD of 5. WHAT-IF post-test. *Ten* of the 41 subjects scored lower on WHAT-IF post-test correctness compared to the pre-test, with a range of 1 to 8 items less correct on the post-test. The average time to complete a WHAT-IF pre-test item was 18 seconds with a SD of 7.4. For the WHAT-IF post-test the average item completion time was 11 seconds, and the SD was 3.1. Four subjects took longer on the post- compared to the pre-test. The range entailed 1.4 to 6.2 seconds for 'extra' average item response time.

With regard to the trade-off effect between correctness and speed no conclusive picture emerged. One statistically significant and positive correlation between WHAT-IF item completion time and WHAT-IF correctness was found. It entailed the correlation computed within subjects across the WHAT-IF post-test items: $r = .31$, $p < .05$. The other correlations computed between answer time and correctness had a value of $r = .25$, $p > .10$, when computed within subjects across the WHAT-IF pre-test items, a value of $r = -.34$, $p = .10$ when computed within WHAT-IF pre-test items across subjects, and finally a value of $r = .05$, $p > .10$ when computed within WHAT-IF post-test items across subjects.

A repeated measurement analysis on the WHAT-IF test scores showed a significant within-subject effect on number of correct items ($F_{1, 39} = 6.98$, $p < .05$).

With respect to the latencies collected, the same picture emerged. A repeated measurement analyses on the WHAT-IF test scores showed a significant within-subject effect of on decrease in item response time ($F_{1, 39} = 45.19$, $p < .001$).

9.5 Relationship between the WHAT-IF test scores and definitional test scores

The definitional knowledge test was given in the same form as pre- and as post-test. It consisted of 20 multiple choice items with 3 alternative answers each. A repeated measurement analysis on the definitional test-scores showed *no* significant within-subject effect of number of correct items ($F_{1, 36} < 1$; due to technical issues the scores of three subjects could not be used). Table 3 displays the correlations between the post-test scores of the WHAT-IF tests and the definitional knowledge test scores.

Table 5. *Correlations between the post-test scores of the WHAT-IF tests and the correctness scores of the definitional knowledge test*

	definitional post-test scores	**WHAT-IF post-test completion times**
WHAT-IF post-test correctness	.41 ($p < .05$)	.31 ($p < .05$)
WHAT-IF post-test completion times	.32 ($p < .05$)	-

Here, we see that moderate and statistically significant correlations exist between the three scores. However, the correlations between the gain in definitional test score and gain in WHAT-IF correctness and time scores are not considerable and do not reach significance levels. The latter findings mean that a gain in intuitive knowledge, either as measured by an increase in WHAT-IF correctness or by a decrease in WHAT-IF item response time, does *not* automatically yield a gain in definitional knowledge.

9.6 Relationship between the WHAT-IF test scores and several process measures

The 41 learners used the support measures to a maximal extent and as a result variation between subjects was small. However, to be able to compare this study with the other two studies we report the correlations between the number of assignments used, the explanations consulted and the post-test scores. Table 4 displays the correlations

Table 6. *Correlations between the WHAT-IF post-test scores and number of assignments and explanations used*

post-test scores	number of assignments	number of explanations
WHAT-IF correctness scores	-.13 ($p > .10$)	-.19 ($p > .10$)
WHAT-IF completion times	-.12 ($p > .10$)	.04 ($p > .10$)

As expected for this study none of the correlations had a considerable size. Therefore, the pattern of the previous studies could not be replicated.

9.7 Discussion

One main notion of this study is that the variance of performance on the post-test measures is considerable. More precisely, a quarter of all the learners had a lower the WHAT-IF post-test correctness score compared to the correctness score of the WHAT-IF pre-test.

Still, both a significant increase in correctness scores and a significant decrease in WHAT-IF item completion times are established, and as in the SETCOM study, *no* increase in definitional knowledge test scores is attained.

Analogous to the Collisions and SETCOM studies, in the CIRCUIT study the correlations between the gain in definitional knowledge and intuitive knowledge are small, which indicate that the tests seem to measure different things and that an increase in intuitive knowledge does not necessarily yield an increase in definitional knowledge.

Finally, as opposed to the previous studies, no interpretable pattern between number of assignments or explanations and post-test scores is found.

10. Conclusion

An overall conclusion from our work is that the tests constructed with WHAT-IF items were able to tap an improvement in learning, both in increases in correctness scores and decreases in item completion times (Studies 1, 2, & 3). The WHAT-IF tests seemed furthermore capable of assessing the positive impact of the use of assignments (Studies 1 & 2), and the less favourable influence of explanations (Study 2). Moreover, there is some evidence in the data that the quicker an item is answered, the higher the chance that it is correctly answered (Studies 1 & 2). These figures correspond to our conceptions on the intuitive quality of knowledge, and to our ideas about which circumstances promote the acquisition of knowledge with an intuitive quality. Therefore, we can say that the validity data are – at a first start – supportive. This conclusion does not exclude the fact that further studies are needed to validate the WHAT-IF item format and make the validity data even more convincing.

As a general conclusion we think that our data show that designing new types of knowledge tests in relation to new types of (technology supported) learning is a promising way to follow. This is in line with studies by, among others, Leutner (1993), Rieber, Boyce, and Assad (1990), Stark, Gruber, Renkl, and Mandl (1995), and White (1993).

References

Berry, D.C., & Broadbent, D.E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, 79, 251-272.

Broadbent, D.E., Fitzgerald, P., & Broadbent, M.H.P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology, 77,* 33-50.

Fischbein, E. (1987). *Intuition in science and mathematics.* Dordrecht, The Netherlands: Reidel.

diSessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction, 10 (2 & 3),* 105-25.

Hayes, N.A., & Broadbent, D.E. (1988). Two modes of learning for interactive tasks. *Cognition, 28,* 249-276.

Jong, T. de, & Joolingen, W.R. van (1995). The SMISLE environment: Learning and design of integrated simulation learning environments. In P. Held & W.F. Kugemann (Eds.) *Telematics for education and training* (pp. 173-187). Amsterdam: IOS Press.

Jong, T., de, & Joolingen, W.R., van (in preparation). *Discovery learning with computer simulations.*

Jong, T. de, Joolingen, W.R., van, Pieters, J.M., van der Hulst, A., & de Hoog, R. (1992). *Instructional support for simulations: overview, criteria and selection.* DELTA project SMISLE, Deliverable D02. University of Twente, Department of Education.

Jong, T.de, Martin, E., Zamarro J-M., Esquembre, F., Swaak, J., & van Joolingen, W.R. (1995, April). *Support for simulation-based learning; the effects of assignments and model progression in learning about collisions.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Joolingen, W.R. van, Hulst, A. van der, Swaak, J, & Jong, T. de (1995). Support for simulation-*based learning; the effects of model progression in learning about oscillations.* SMISLE project, deliverable D24a. Enschede: University of Twente.

Leutner, D. (1993). Guided discovery learning with computer-based simulation games: effects of adaptive and non-adaptive instructional support. *Learning and Instruction, 3,* 113-132.

Norman, D. A. (1993). *Things that make us smart: defending human attributes in the age of the machine.* New York: Voyager.

Rieber, L.P., Boyce, M., & Assad, C. (1990). The effects of computer animation on adult learning and retrieval tasks. *Journal of Computer-Based Instruction, 17,* 46-52.

Snow, R. E. (1989). Toward assessment of cognitive and conative structures in learning. *The Educational Researcher, 18,* 8-14.

Stark, R., Gruber, H., Renkl, A., & Mandl, H. (1995). *Dissociation and convergence of instructional effects in complex learning.* Research report no. 63. Institute of Educational Psychology, University of Munich.

Swaak, J., & Jong, T. de (1996). Measuring intuitive knowledge in science: the development of the WHAT-IF test. *Studies in Educational Evaluation, 22 (4),* 341-362.

Swaak, J., Blokhuis, V., Gutierrez, M., & López, M. (in preparation). SERVIVE; initial learner evaluation studies. SERVIVE project, deliverable D3.2. Enschede: University of Twente.

Thomas, R., & Hooper, E. (1991). Simulations: An opportunity we are missing. *Journal of Research on Computing in Education, 23(4),* 497-513.

Westcott, M.R. (1968). *Towards a contemporary psychology of intuition.* New York: Holt, Rinehart & Winston.

White, B.Y. (1993). ThinkerTools: causal models, conceptual change, and science education. *Cognition and Instruction, 10,* 1-100.

White, B.Y., & Frederiksen, J.R. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence, 42,* 99-157.

Acknowledgements

The work presented here is carried out under the DELTA programme of the EC as projects D2007 and ET 1020. We gratefully acknowledge the contributions of all my present and former colleagues of the SMISLE and the SERVIVE projects. Furthermore, we specially would like to express our gratitude to Ernesto Martin, Jose-Miguel Zamarro, and Fransisco Esquembre (University of Murcia, Spain), Wouter R. van Joolingen, Mark Gellevij, and Koen Veermans (University of Twente, the Netherlands), Anja van der Hulst and Robert de Hoog (University of Amsterdam, the Netherlands), Vincent Blokhuis (Twents MBO College, the Netherlands) for their contribution in constructing the tests and/or conducting the experiments described in this paper.